How Contexts Matter Understanding in Dialogues

YUN-NUNG (VIVIAN) CHEN

國立臺灣大學
National Taiwan University

HTTP://VIVIANCHEN.IDV.TW

Outline

- Word-Level Contexts in Sentences
  - Learning from Prior Knowledge –

    Knowledge-Guided **S**tructural **A**ttention **N**etworks (K-SAN) [Chen et al., '16]
  - Learning from Observations –

    Modularizing **U**nsupervised **S**ense **E**mbedding (MUSE) [Lee & Chen, '17]
- Sentence-Level Contexts in Dialogues
  - Investigation of Understanding Impact –

    Reinforcement Learning Based Neural Dialogue System [Li et al., '17]
- Conclusion

Task-Oriented Dialogue System

- Dialogue systems are intelligent agents that are able to help users finish tasks more efficiently via conversational interactions.
- Dialogue systems are being incorporated into various devices (smart-phones, smart TVs, in-car navigating system, etc).

JARVIS – Iron Man's Personal Assistant

Baymax – Personal Healthcare Companion

Context in Language

- Word-level context
  - Prior knowledge such as linguistic syntax

    show me the flights from seattle to san francisco

  - Collocated words

    Smartphone companies including apple, blackberry, and sony will be invited.

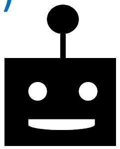
Contexts provide informative cues for better understanding

- Sentence-level context

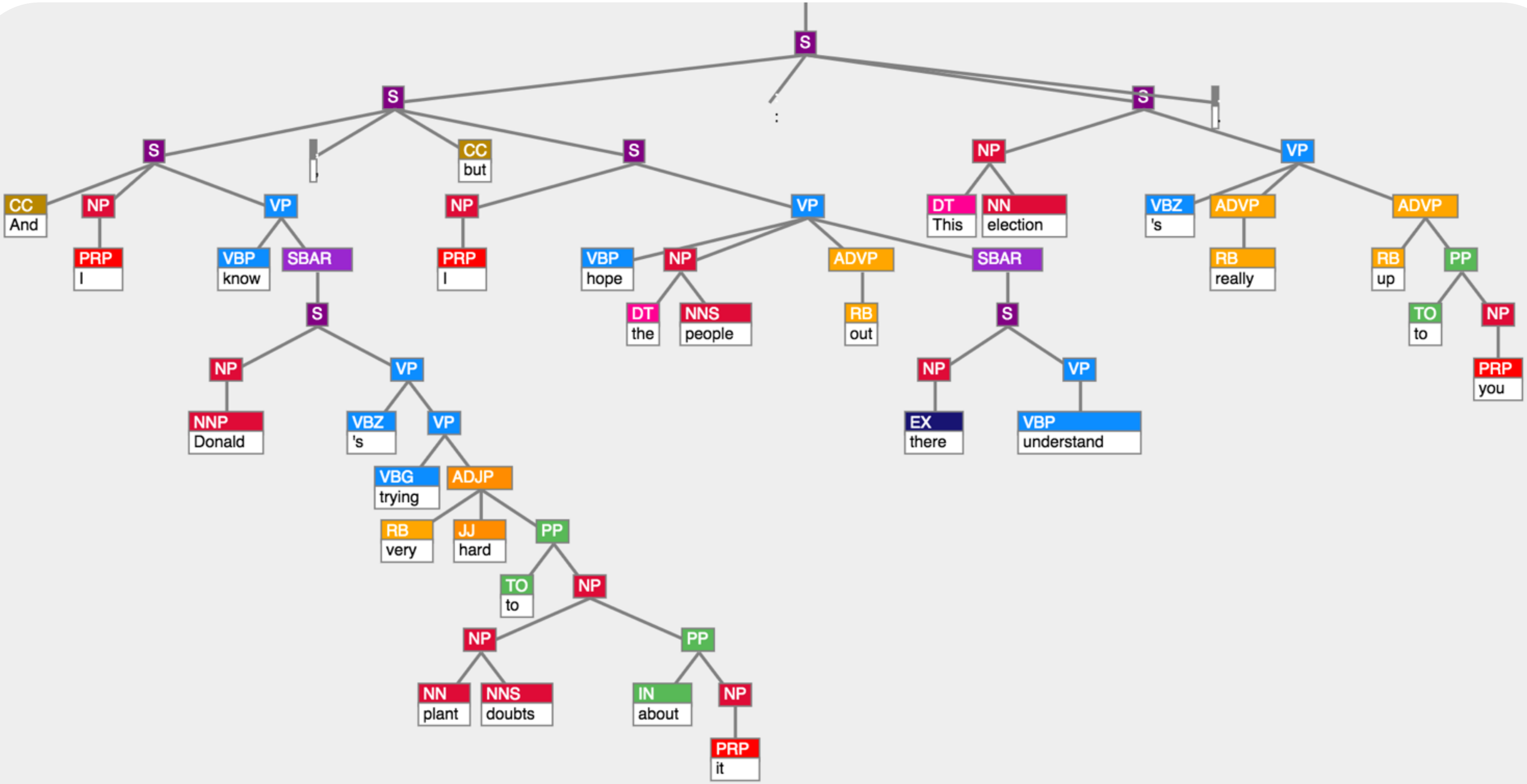
(browsing action movie reviews…)
Find me a good one this weekend

request_movie
(genre=action, date=this weekend)

London Has Fallen is currently the number 1 action movie in America



How misunderstanding influences the dialogue system performance



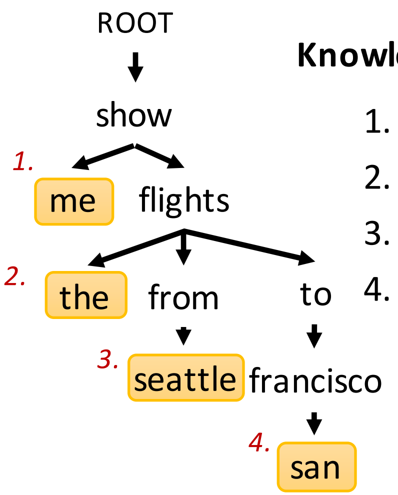
Knowledge-Guided Structural Attention Network (K-SAN)

Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

Sentence Structural Knowledge

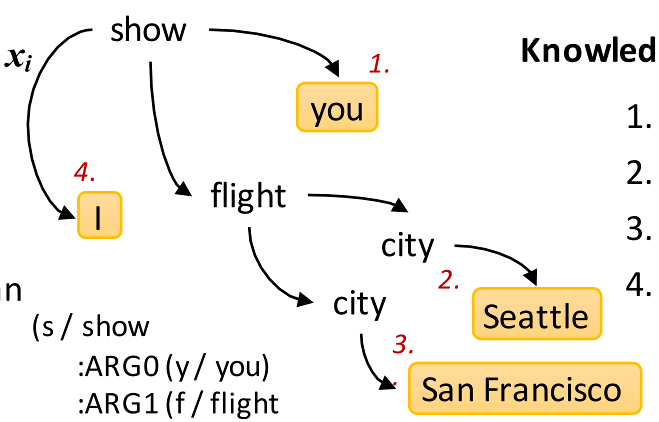
- Syntax (Dependency Tree)
- Semantics (AMR Graph)

Sentence s show me the flights from seattle to san francisco



Knowledge-Guided Substructure x_i

1. show me
2. show flights the
3. show flights from seattle
4. show flights to francisco san



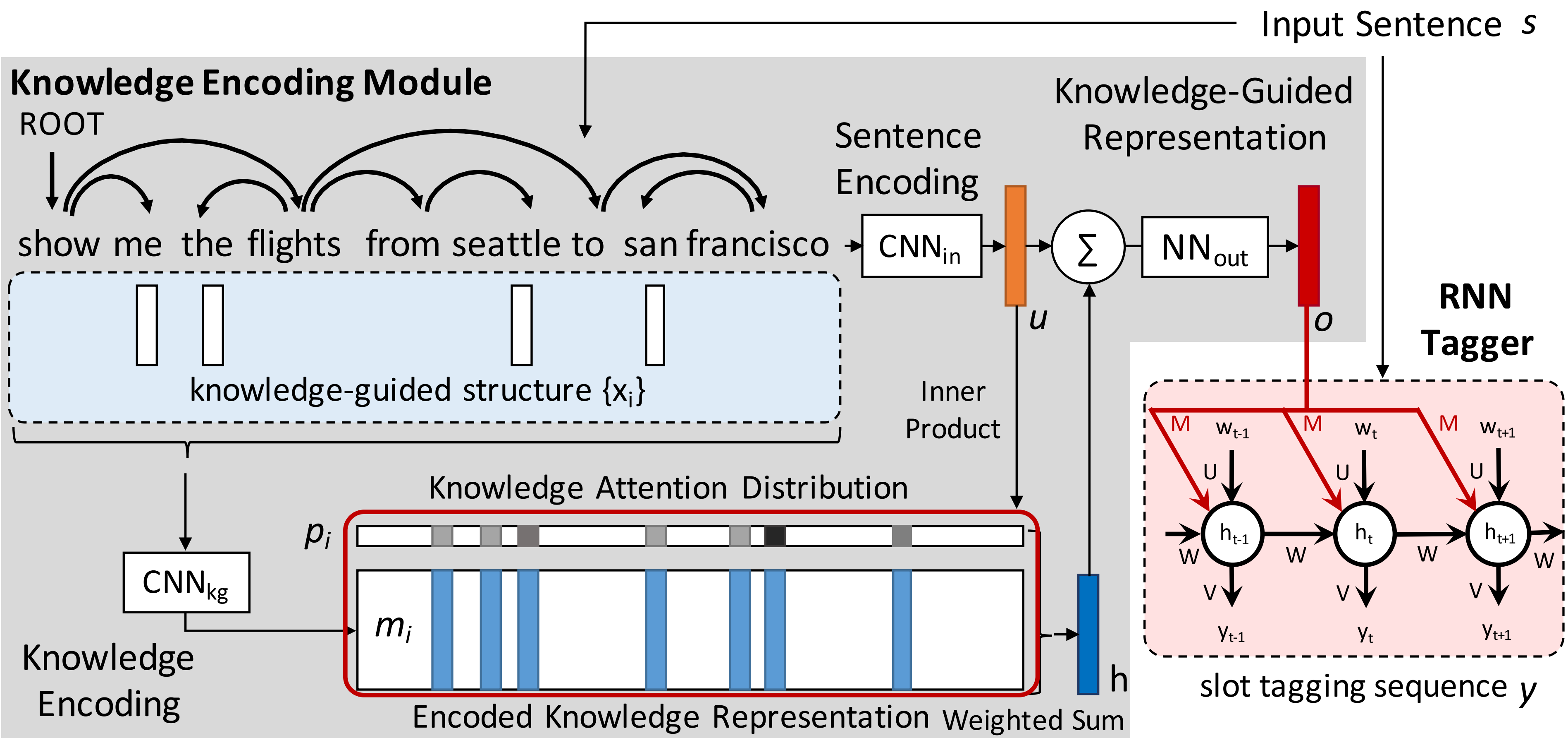

```
(s / show
    :ARG0 (y / you)
    :ARG1 (f / flight
        :source (c / city
            :name (d / name :op1 Seattle))
        :destination (c2 / city
            :name (s2 / name :op1 San :op2 Francisco)))
    :ARG2 (i / I)
    :mode imperative)
```

Knowledge-Guided Substructure x_i

1. show you
2. show flight seattle
3. show flight san francisco
4. show i

Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

Knowledge-Guided Structures

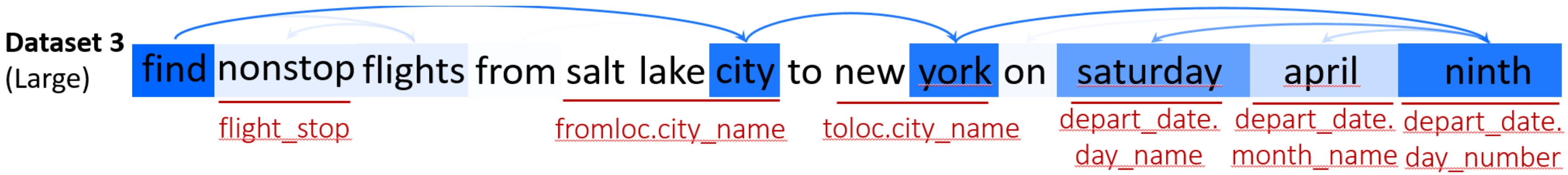


The model will pay more attention to more important substructures that may be crucial for slot tagging.

Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, “Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks,” preprint arXiv: 1609.00777, 2016.

Attention Analysis

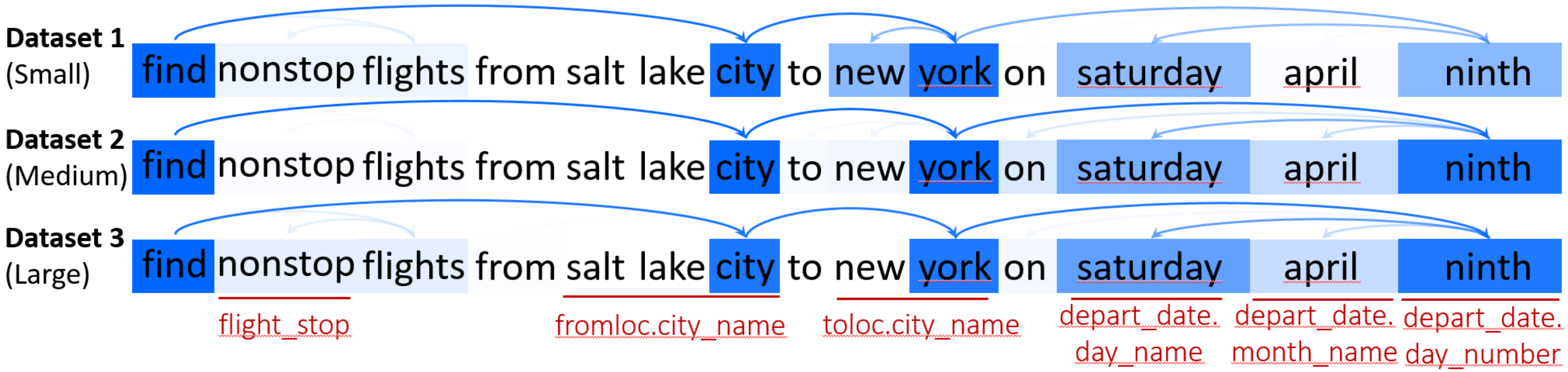
- Darker blocks and lines correspond to higher attention weights



Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

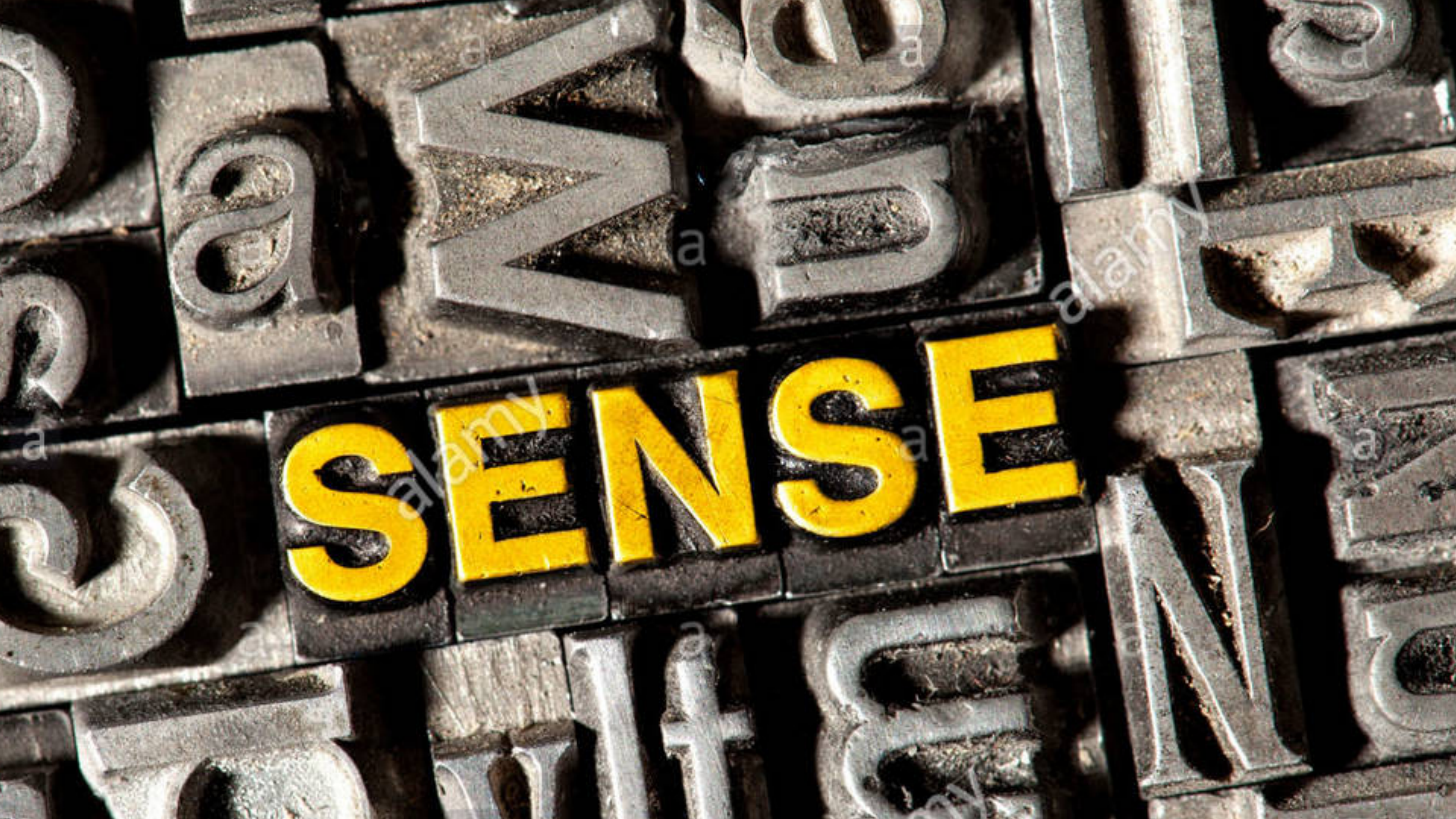
Attention Analysis

- Darker blocks and lines correspond to higher attention weights



K-SAN learns the similar attention to salient substructures with less training data

Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

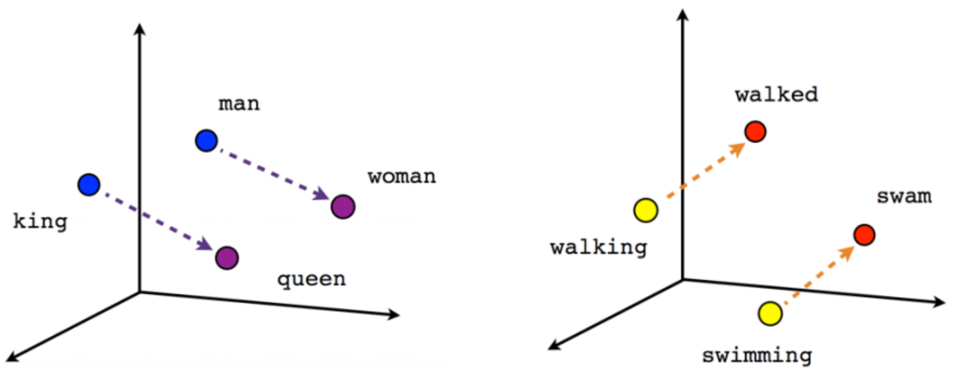


Modularizing Unsupervised Sense Embeddings (MUSE)

G.-H. Lee and Y.-N. Chen, "MUSE: Modularizing Unsupervised Sense Embeddings," in *EMNLP*, 2017.

Word Embedding

- Word embeddings are trained on a corpus in an unsupervised manner



Finally I chose Google instead of <u>Apple</u>.

Can you buy me a bag of <u>apples</u>, oranges, and bananas?

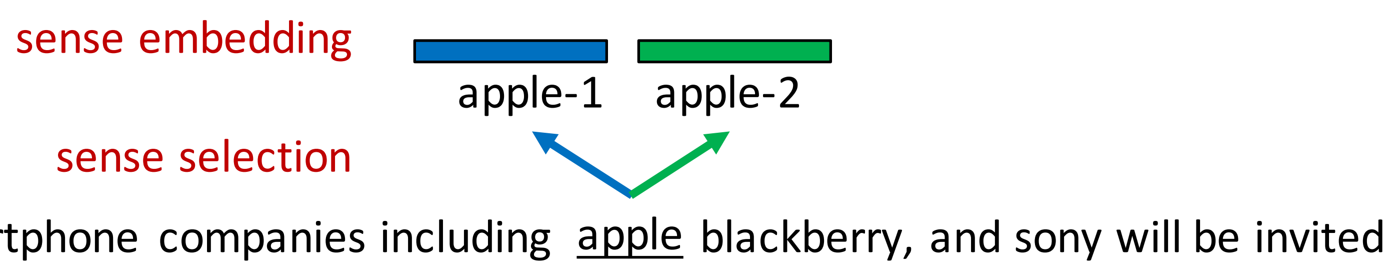
- Using the **same embeddings** for **different senses** for NLP tasks, e.g. NLU, POS tagging

Words with different senses should correspond different embeddings

G.-H. Lee and Y.-N. Chen, "MUSE: Modularizing Unsupervised Sense Embeddings," in *EMNLP*, 2017.

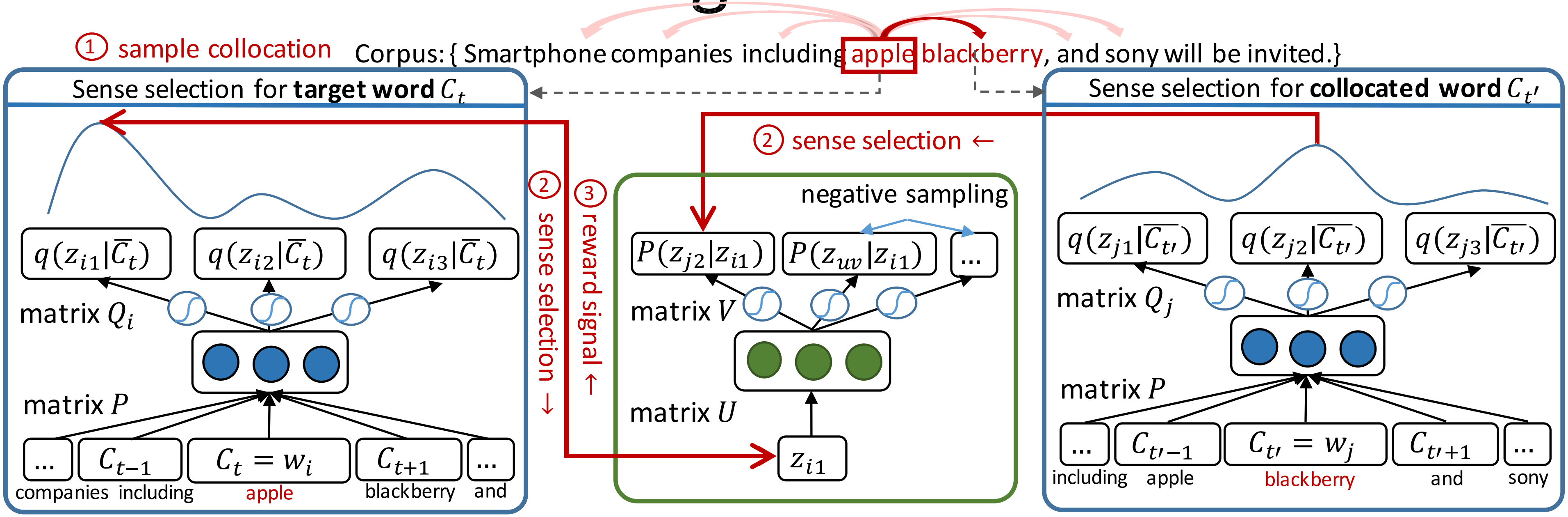
Task – Unsupervised Sense Embeddings

- Input: unannotated text corpus
- Two key mechanisms
  - **Sense selection** given a text context
  - **Sense representation** to embed statistical characteristics of sense identity



G.-H. Lee and Y.-N. Chen, "MUSE: Modularizing Unsupervised Sense Embeddings," in *EMNLP*, 2017.

MUSE: Modularizing Unsupervised Sense Embeddings



Sense Selection Module **Sense Representation Module** **Sense Selection Module**

- Sense selection
  - Policy-based

$$\pi(z_{ik} \mid \bar{C}_t) = \frac{\exp(Q_{ik}^T \sum_{j \in \bar{C}_t} P_j)}{\sum_{k' \in Z_i} \exp(Q_{ik'}^T \sum_{j \in \bar{C}_t} P_j)}$$

  - Value-based

$$q(z_{ik} \mid \bar{C}_t) = \sigma(Q_{ik}^T \sum_{j \in \bar{C}_t} P_j)$$

- Sense representation learning

$$\log \mathcal{L}(z_{jl} \mid z_{ik}) = \log \frac{\exp(U_{z_{ik}}^T V_{z_{jl}})}{\sum_{z_{uv}} \exp(U_{z_{ik}}^T V_{z_{uv}})}$$

  - Skip-gram approximation

$$\log \bar{\mathcal{L}}(z_{jl} \mid z_{ik}) = \log \sigma(U_{z_{ik}}^T V_{z_{jl}}) + \sum^{M} \mathbb{E}_{z_{uv} \sim p_{neg}(z)}[\log \sigma(-U_{z_{ik}}^T V_{z_{uv}})]$$

Collocated likelihood serves as a reward signal to optimize the sense selection module.

Contextual Word Similarity Experiments

- Dataset: SCWS for multi-sense embedding evaluation

He borrowed the money from **banks**.	I live near to a **river**.	correlation=?

Approach	MaxSimC	AvgSimC
Huang et al., 2012	26.1	65.7
Neelakantan et al., 2014	60.1	**69.3**
Tian et al., 2014	63.6	65.4
Li & Jurafsky, 2015	66.6	66.8
Bartunov et al., 2016	53.8	61.2
Qiu et al., 2016	64.9	66.1
MUSE-Policy	66.1	67.4
MUSE-Greedy	66.3	68.3
MUSE-ε-Greedy	**67.4**[+]	68.6

Qualitative Analysis

Context	… braves finish the season in **tie** with the los angeles dodgers …	… his later years proudly wore **tie** with the chinese characters for …
k-NN	scoreless otl shootout 6-6 hingis 3-3 7-7 0-0	pants trousers shirt juventus blazer socks anfield
Figure		

Qualitative Analysis

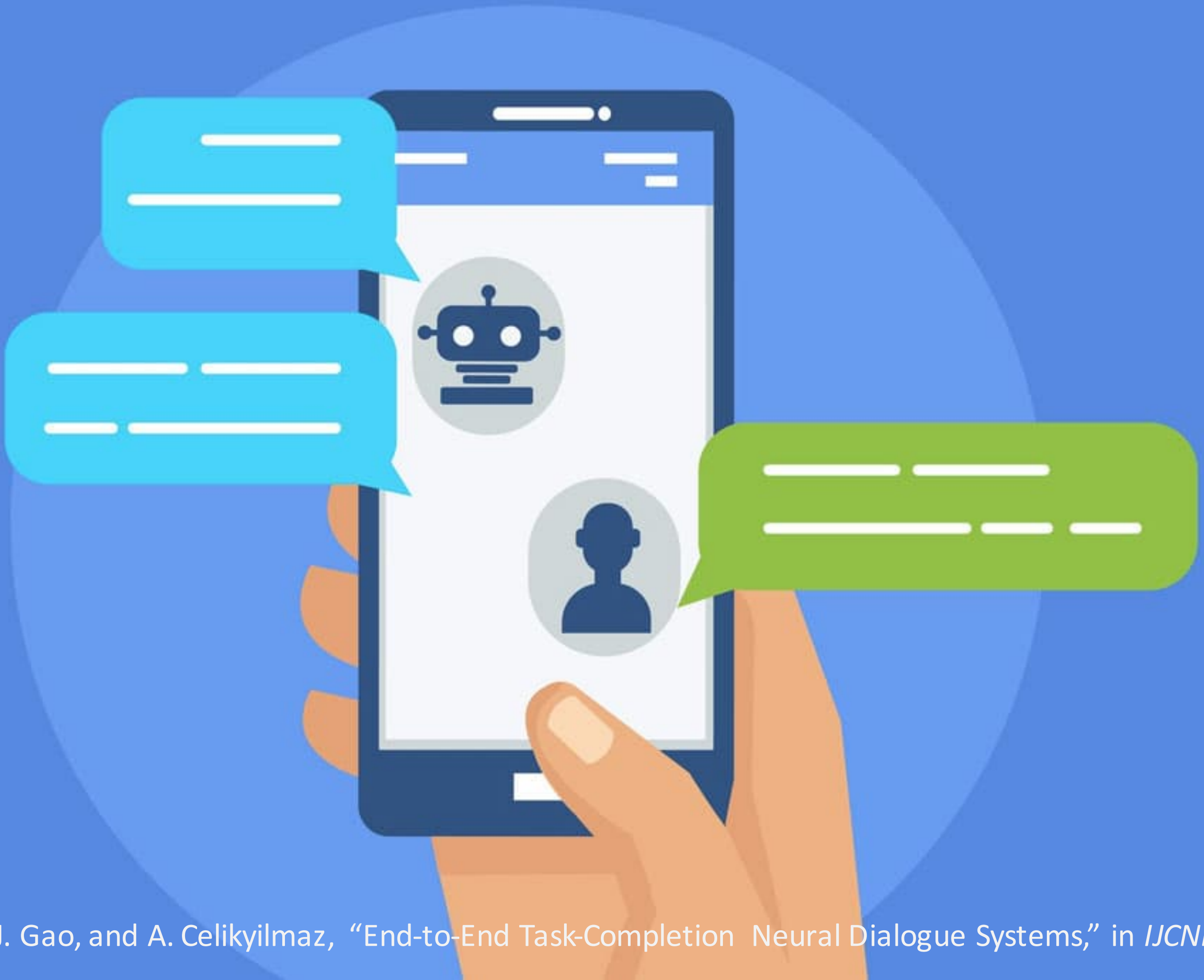
Context	… of the mulberry or the **blackberry** and minos sent him to …	… of the large number of **blackberry** users in the us federal …
k-NN	cranberries maple vaccinium apricot apple	smartphones sap microsoft ipv6 smartphone
Figure		

Qualitative Analysis

Context	… shells and/or high explosive squash **head** and/or anti-tank …	… head was shaven to prevent **head** lice serious threat back then …	… appoint john pope republican as **head** of the new army of …
k-NN	venter thorax neck spear millimeters fusiform	shaved thatcher loki thorax mao luther chest	multi-party appoints unicameral beria appointed
Figure			

MUSE learns sense embeddings in an ***unsupervised*** way and achieves the first ***purely sense-level*** representation learning system with ***linear-time sense selection***

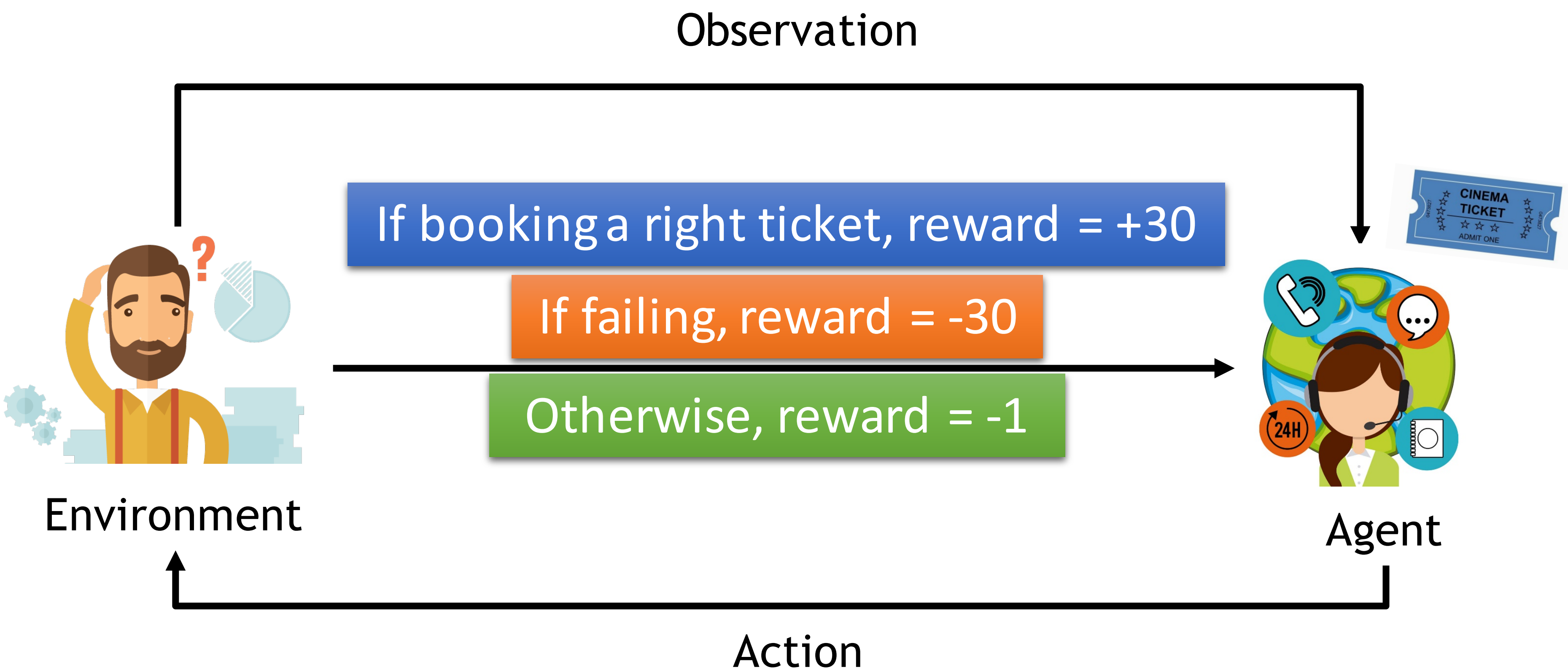
RL-Based Neural Dialogue Systems



X. Li, Y.-N. Chen, L. Li, J. Gao, and A. Celikyilmaz, "End-to-End Task-Completion Neural Dialogue Systems," in *IJCNLP*, 2017.

E2E Neural Dialogue System

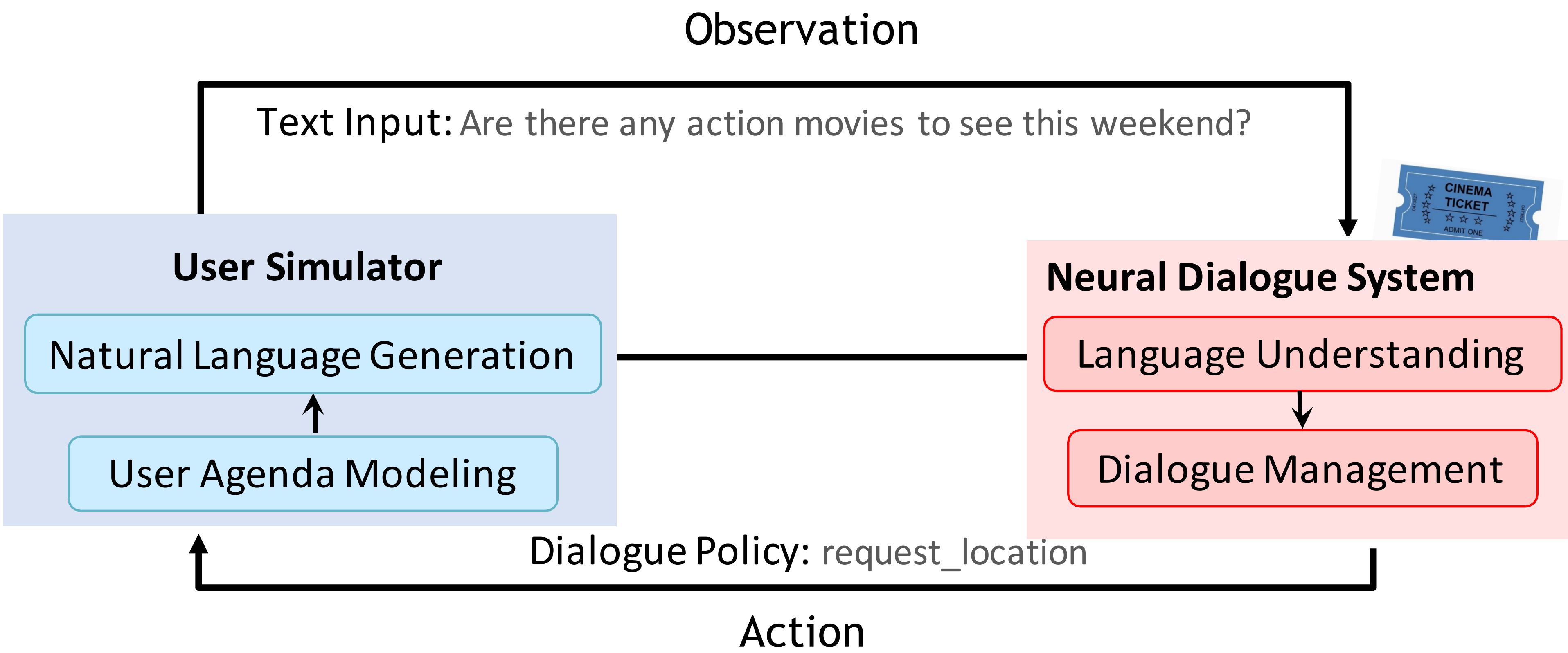
- **Dialogue management** is framed as a **reinforcement learning** task
- Agent learns to select actions to **maximize the expected reward**



X. Li, Y.-N. Chen, L. Li, J. Gao, and A. Celikyilmaz, "End-to-End Task-Completion Neural Dialogue Systems," in *IJCNLP*, 2017.

E2E Neural Dialogue System

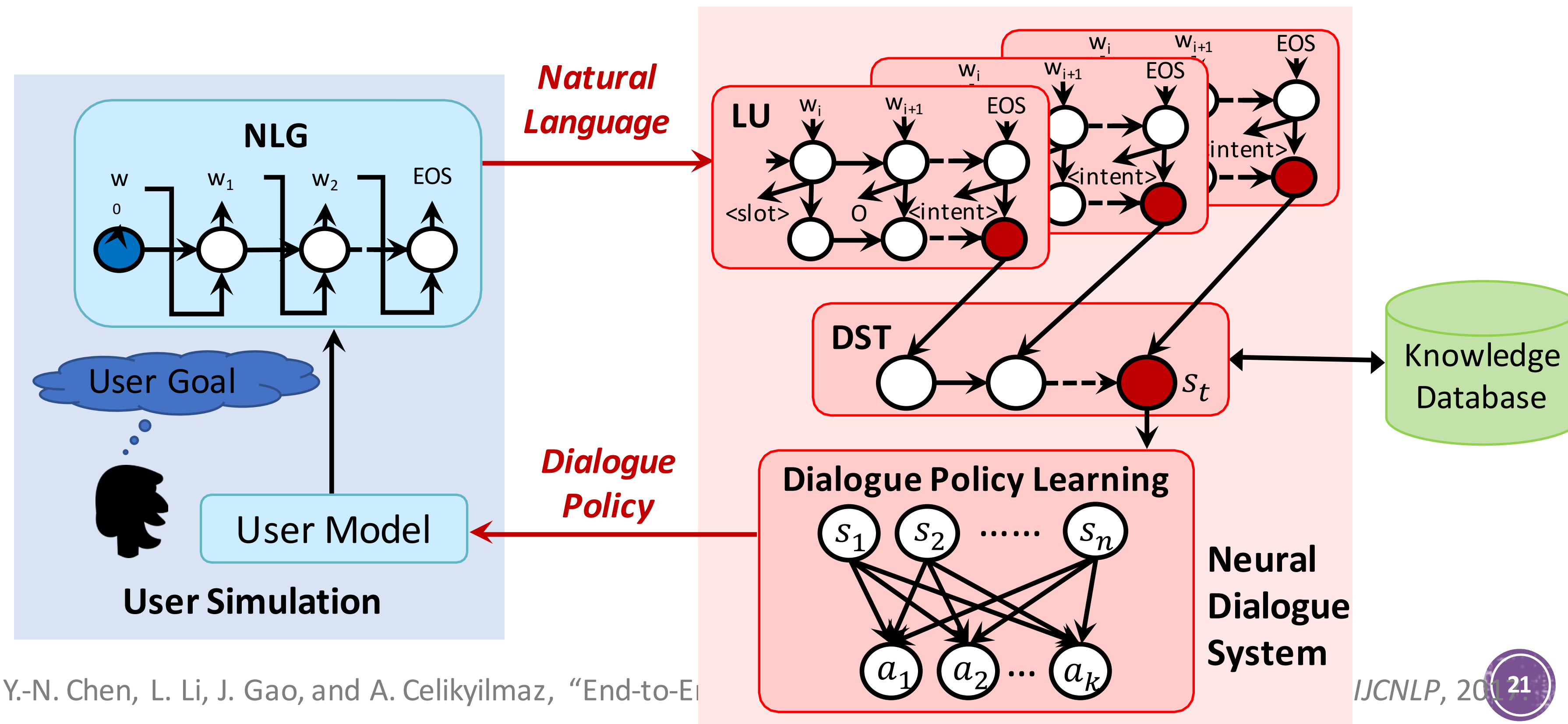
- **Dialogue management** is framed as a **reinforcement learning** task
- Agent learns to select actions to **maximize the expected reward**



X. Li, Y.-N. Chen, L. Li, J. Gao, and A. Celikyilmaz, "End-to-End Task-Completion Neural Dialogue Systems," in *IJCNLP*, 2017.

End-to-End Neural Dialogue Systems

- LU, DST (neural dialogue system), and NLG (user simulation) are trained in supervised way
- End-to-end training for dialogue policy learning



X. Li, Y.-N. Chen, L. Li, J. Gao, and A. Celikyilmaz, “End-to-E… *IJCNLP*, 201…

Dialogue Examples

User goal – ***Two tickets for "Deadpool" tomorrow 9PM at AMC Pacific Place 11 theater, Seattle.***

RULE BASED SYSTEM

User: Do you have any ticket for movie 'Deadpool' at 9:00 pm?
Agent: Which theater do you want?
User: AMC pacific place 11 theater, please!
Agent: How many tickets do you want?
User: Two!
Agent: What date?
User: Tomorrow!
Agent: Which city do you want?
User: Seattle!
Agent: Do you have any specification for zip code?
User: No, I do not care.
Agent: Do you have any distance constraint?
User: No, I do not care.
Agent: How about video format? 2d? Imax 3d?
User: Any kind is okay.
Agent: Do you have any constraint to price?
User: No!
Agent: Okay, I will issue 2 tickets for you, tomorrow 9:00 pm at AMC pacific place 11 theater, Seattle, movie 'Deadpool'.
User: Okay, thanks.
Agent: Thanks!

Issue 1: no notion about what requests can be skipped

REINFORCEMENT LEARNING SYSTEM

User: I want two tickets for movie 'Deadpool'.
Agent: What is the starttime you prefer?
User: 9:00 PM
Agent: Which day?
User: Tomorrow
Agent: Which city?
User: Seattle
Agent: Which theater do you prefer?
User: AMC Pacific Place 11 theater.
Agent: Okay, I will issue 2 tickets for you, tomorrow 9:00 pm at AMC pacific place 11 theater, Seattle, movie 'Deadpool'.
User: Okay, thanks.
Agent: Thanks!

Skip the requests the user may not care about to improve efficiency

Dialogue Examples

User goal – ***Two tickets for "Hail Caesar" tomorrow 7:15PM at AMC Lowes Oak tree 6, Seattle.***

RULE BASED SYSTEM

User: Do you any ticket for movie 'Hail Caesar' at theater AMC Lowes Oak Tree 6?
Agent: What start time do you prefer?
User: **7:15 pm** Tomorrow
Agent: Which city you are looking for?
User: Seattle
Agent: How many tickets do you want?
User: Two
Agent: Okay, I have issued two tickets for movie 'Hail Caeser' tomorrow **7:50 pm** at theater AMC Lowes Oak Tree 6, Seattle.
User: Oh, no.
Agent: Thanks!

Issue 2: no principle way to deal with LU errors/uncertainty

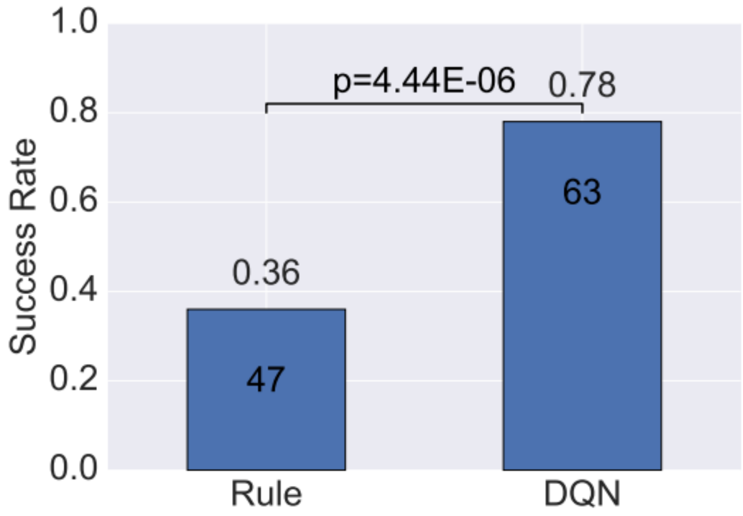
REINFORCEMENT LEARNING SYSTEM

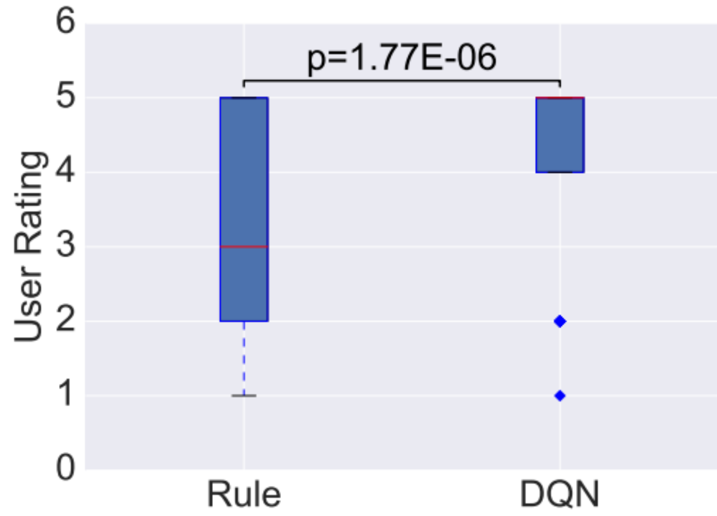
Agent: What start time do you prefer?
User: **7:15 pm** Tomorrow
Agent: Which city you are looking for?
User: Seattle
Agent: Okay, I will issue you tickets for movie 'Hail Caeser' tomorrow **7:50 pm** at theater AMC Lowes Oak Tree 6, Seattle.
User: Oh, no.
Agent: **It is 7:50 pm, right?**
User: Two tickets for 7:15 pm, please.
Agent: Okay, I will issue you two tickets for movie 'Hail Caeser' tomorrow **7:15 pm** at theater AMC Lowes Oak Tree 6, Seattle.
User: Thanks!
Agent: Thanks!

Automatically learn when and what to confirm for better robustness

Human Evaluation

- Users presented with a predefined user goal
- User rating: 1 (worst) to 5 (best) based on both *naturalness* and *coherence* of the dialogue





Conclusion

- Word-level contexts in sentences help understand word meanings
  - Learning from Prior Knowledge –

    K-SAN achieves better LU via known knowledge [Chen et al., '16]
  - Learning from Observations –

    MUSE learns sense embeddings with efficient sense selection [Lee & Chen, '17]
- Sentence-level contexts have different impacts on dialogue performance
  - Investigation of Understanding Impact –

    Slot errors degrade system performance more than intent errors [Li et al., '17]
- Contexts from different levels provide cues for better understanding in supervised and unsupervised ways

Thanks for Attention!